

# Investigating sound and structure in concert: A pupillometry study of relative clause attachment

Jesse A. Harris (jharris@humnet.ucla.edu)

UCLA Department of Linguistics, 3125 Campbell Hall  
Los Angeles, CA 90095 USA

Alexandra Lawn (alawn@ucla.edu)

UCLA Department of Spanish and Portuguese, 5310 Rolfe Hall  
Los Angeles, CA 90095 USA

Marju Kaps (mkaps@ucla.edu)

UCLA Department of Linguistics, 3125 Campbell Hall  
Los Angeles, CA 90095 USA

## Abstract

Listeners must integrate multiple sources of information to construct an interpretation of a sentence. We concentrate here on the alignment of prosodic and syntactic grouping during online sentence comprehension. We present the results from a pupillometry study on the use of prosodic boundaries in resolving well-known attachment ambiguities. Using growth curve analyses to capture the non-linear dynamics of pupil dilation, we found increased pupil excursions for sentences that were disambiguated towards the dispreferred, non-local relation, especially when accompanied by supporting prosodic information. However, when prosodic and structural information did not align, pupillary response was muted, potentially indicating a failure to commit to a specific interpretation. More generally, the study shows how the currently under-utilized pupillometry method offers insights into spoken language comprehension.

**Keywords:** Relative clause attachment; prosody; pupillometry

## Introduction

Sentences like (1) are structurally ambiguous. The relative clause (RC) after a complex noun phrase (NP) may be interpreted as modifying the first noun (NP1; in *High attachment*: the maid was on the balcony) or the second noun (NP2; in *Low attachment*: the actress was on the balcony).

- (1) Someone shot the maid<sub>NP1</sub> of the actress<sub>NP2</sub>  
[RC who was on the balcony]

Many classical theories of sentence processing assume a strong role for relations privileging locality (Kimball, 1973; Frazier & Fodor, 1978) or recency (Gibson, 1998). For example, the principle of *Late closure* prompts the language parser to resolve ambiguous strings like (1) towards a structure in which the RC attaches to the most recently accessed constituent that is grammatically possible, resulting in a Low attachment interpretation. Under their strongest interpretation, locality or recency based principles represent universal properties of the human language processing system and are not subject to variation within languages or individuals.

However, subsequent research has argued that the preferred interpretations of strings like (1) are moderated by a great many other factors, such as plausibility, experience, or prosodic grouping, and may even reflect parsing preferences from specific languages (see Fernández, 2003 for review).

Several explanations for this variation have been proposed. Perhaps the first study to find a High attachment preference for non-local RC modification was Cuetos & Mitchell (1988), who proposed that the statistics of a language strongly influence how the processor resolves ambiguity – i.e., a language shows a High attachment preference because this resolution is the most frequent in the language. Others have explained the variation in terms of additional constraints competing with a universal recency bias (Gibson et al., 1996), or with respect to the availability of alternative parses in different languages (Hemforth et al., 2000; Grillo & Costa, 2014; Grillo et al., 2015). Others still have exempted relative clauses from the domain of Late closure, instead arguing that they are construed using a collection of grammatical and non-grammatical interpretive principles (Gilboy et al., 1995; Frazier & Clifton, 1996). While many factors may well contribute to relative clause attachment preferences, we focus here on the relation between prosodic and syntactic grouping during online sentence comprehension.

In general, prosody refers to the organizational structure of speech, expressed through changes in pitch, duration, and amplitude, among other factors (Ladd, 2008). Although the prosodic grouping of words can reflect metrical preferences (such as the location of prominence or the alternation of feet), it can also signal higher levels of linguistic structure, especially syntactic or information structural relations. We adopt a simplified description of prosodic information, and concentrate on how prosodic groups are formed via the presence of a prosodic boundary (%), which can be indicated by word final lengthening, pitch movements that conventionally mark the end of a group, and pauses in the speech signal.

Previous work indicates that the location of a prosodic boundary between the complex NP and the relative clause disambiguates RC attachment. A prosodic boundary placed between NP2 and the RC (2b) results in a bias towards High attachment, a generalization that appears to be robust across languages (Jun, 2003). In contrast, a boundary separating NP1 from NP2 (2a) appears to reinforce the low attachment construal, even in languages with a general high attachment preference (Fromont et al., 2017). We assume that in such cases comprehenders interpret the prosodic boundary loca-

tion as a cue to syntactic constituency (depicted as parentheses below) in an attempt to align prosodic and syntactic junctures.

- (2) Someone shot ...
- a. (the maid) % (of the actress who was on the balcony)
  - b. (the maid of the actress) % ( who was on the balcony)

As most studies on attachment ambiguity in complex NPs are conducted using offline measures (though see, for instance, Kim & Christianson, 2013 and Fernández & Sekerina, 2015), it is unclear whether prosodic boundaries guide online sentence processing, and if so whether each boundary location is used immediately. We report the results of a pupillometry study addressing the role of prosodic boundaries on attachment ambiguity resolution in real time comprehension. The experiment was designed to explore two distinct possibilities. First, it is possible that non-local dependencies are computationally taxing to process, regardless of prosodic boundary information. In this case, we would predict that (i) sentences grammatically disambiguated to a high attachment construal would elicit processing costs, which (ii) would not be mitigated by corroborating prosodic information. As an alternative, it is possible that non-local dependencies are avoided for independent prosodic reasons, e.g., if a boundary after NP1 is preferred to keep the prosodic units of equal weight (Fodor, 1998). Such a view would predict that high attachment resolutions are only costly when not supported by prosodic boundary information.

### Pupillometry

Pupil dilation is likely to reflect a multitude of components, some related to demands on cognition and attention, and others to the autonomic nervous system. The pupil naturally fluctuates, dilating in response not only to neural inhibition from the parasympathetic oculomotor system and the noradrenergic system, but also to the presentation of an external stimulus (Wilhelm et al., 1999), including emotionally arousing stimuli, challenging math problems, and unconscious or barely discernable stimuli (Beatty & Kahneman, 1966; Kahneman & Beatty, 1966; Hess & Polt, 1960; Laeng et al., 2012).

Although the tools and techniques for pupillometry are still developing, existing literature has already provided convincing evidence that pupil dilation indirectly indexes increased demands on mental effort and attention associated with difficult to interpret material along various linguistic dimensions (Schmidtke, 2018 for review). While some early studies found a relation between increased pupil dilation and syntactic complexity (Just & Carpenter, 1993), more recent studies have begun to explore a wider range of ways in which pupil size might reflect other factors in language comprehension. Major findings include an association between increased pupil size and structurally complex sentences (Demberg & Sayeed, 2016), prosodic disambiguation in garden path sentences (Engelhardt et al., 2010), semantic anomalies

(Demberg & Sayeed, 2016), lexical frequency or increased emotional valence (Kuchinke et al., 2007), violations of expected meter (Scheepers et al., 2013), and inadequate or misleading pitch accent (Zellin et al., 2011). In keeping with the current pupillometry literature, we will assume, as a basic linking hypothesis between cognition and behavior, that increased pupil size indirectly reflects increased cognitive load, including mental effort directed at managing language comprehension processes.

Further, pupillometry offers an appealing supplement to other online methods. It is easy to administer and comparatively inexpensive, while offering an online and passive measurement. As pupil size is not under conscious control, pupil dilation measurements are likely to be resilient to task-specific strategies that subjects may learn or employ during the experiment. Thus, pupillometry studies offer a promising avenue for exploring the role of prosodic information in resolving structural ambiguity.

## Pupillometry study

### Participants

Forty-eight native college-aged speakers of American English were recruited from a Psychology Subject Pool at the University of California, Los Angeles, and were compensated with course credit. No participant reported any history of hearing loss or language disorder. Experimental sessions typically lasted no more than 30 minutes.

### Materials and method

Twenty quartets were constructed in a 2×2 design. Quartets crossed Boundary location (post-NP1, post-NP2) and Attachment (High, Low). All sentences involved a complex noun phrase (*the brother of the musicians*) containing two noun phrases (NP1, NP2) of opposite grammatical number, followed by a relative clause disambiguated to high (modifying NP1: *the brother*) or low (modifying NP2: *the musicians*) attachment. The attachment height was grammatically specified by the plurality of the verb (*was, were*) within the relative clause (*who was really quiet*), which was kept constant within each quartet. Half of the items were disambiguated by the singular form of the auxiliary (*was*), half by the plural form (*were*). A sample item is shown in Table 1. In addition, comprehension questions were presented after half of the items to encourage participants to pay attention. Questions did not ask about relative clause attachment height.

We obtained measures for the lexical level characteristics of length, frequency and number of syllables using the English Lexicon Project (Balota et al., 2007) for NP1 and NP2. Nouns did not differ on length, number of syllables, or frequency, as determined by several measures, including (log) HAL (Lund & Burgess, 1996) and (log) SUBTLEX<sub>US</sub> (Brysbart & New, 2009).

Sentences were recorded with a high-quality microphone in a sound attenuated chamber by a trained phonologist. Audio files were truncated after the relative clause (marked by

Boundary	High attachment	Low attachment
NP1	Everybody met the <u>brother</u> % of the musicians who was really quiet // although the club was really noisy.	Everybody met the brothers % of the <u>musician</u> who was really quiet // although the club was really noisy.
NP2	Everybody met the <u>brother</u> of the musicians % who was really quiet // although the club was really noisy.	Everybody met the brothers of the <u>musician</u> % who was really quiet // although the club was really noisy.

Table 1: Sample quartet item crossing Boundary location (NP1, NP2) and Attachment (High, Low). The underlining identifies the noun modified by the relative clause (*who was really quiet*). The prosodic boundary is indicated by the % symbol.

the // symbol in Table 2). 100ms of computer generated silence was inserted after the relative clause, and the post-relative clause segment was spliced into the recording, so that pupillary response was recorded on acoustically identical material within items. Items were then acoustically normalized to make the transition into the spliced segment as seamless as possible.

The 20 experimental item quartets were presented in counterbalanced and individually randomized order, and were interspersed with 40 items from two separate experiments (one also manipulating boundary placement, and another manipulating contrastive accent), along with 26 filler items unrelated to any experiment. Participants were instructed to fixate on a cross at the center of the screen without blinking for the duration of the sentence. They were encouraged to blink before and after the sentence presentation, and to rest their eyes as needed, in order to minimize the possibility of eye blinks due to fatigue.

Items were presented with Experiment Builder (SR Research) to subjects over sound-isolating headphones in a moderately lit room dedicated to experimentation. Eye position and pupil area were recorded using an SR Research EyeLink 1000 Plus eye tracker sampling at 500Hz. The tracker was mounted to the table at 55cm from a 27 inch LCD monitor with a light gray background. A 5-point calibration procedure was used before recording and as necessary, and drift correction was conducted between every trial.

## Results and discussion

Subjects performed very well on post-sentence comprehension questions ( $M = 96\%$ ), which did not probe the interpretation of the relative clause. There was a marginal effect of mismatches between Boundary location and Attachment, so that comprehension questions following High attachment sentences were less accurate when paired with an NP2 boundary location ( $diff = 3\%$ ), and questions after Low attachment sentences were less accurate when paired with an NP1 boundary ( $diff = 3\%$ ),  $t = 1.96, p = 0.05$ . No other effects were observed. The pattern suggests that incompatible boundaries interfered with comprehension, but that there was no general effect of boundary or attachment on general comprehension. However, performance on all conditions was uniformly high, averaging at 94% or above.

Pre-processing of gaze location and pupil size was conducted in Data Viewer. Fixations outside of the fixation cross were removed. Trials with eye blinks (less than 5% of the to-

tal data) were also removed to avoid reconstructing pupil size during noisy trials. Mean pupil size was downsampled into 100 20ms bins starting from the end of the relative clause, for a total recording time of 2000ms past the relative clause. The remaining analyses were conducted in R (R Core Team, 2016). Data were fit with a growth curve model (Mirman, 2016) to avoid assuming a linear form or an arbitrary time window for analysis. Growth curve models have been used previously to quantify continuous changes in pupillary response, and we adopt those authors' interpretations of the curve with respect to pupil response (Kuchinsky et al., 2013; McGarrigle et al., 2017).

We report a third-order (cubic) orthogonal polynomial model with fixed effects of Boundary, Attachment, and their interaction on polynomial terms, and by-subject and by-item random slopes (Baayen et al., 2008), as shown in Table 2. Experimental predictor variables received deviation (sum) coding with NP1 and Low Attachment conditions specified as reference levels for their respective factors. Orthogonal polynomials were used to avoid extreme multicollinearity between adjacent samples in the time series.

	Estimate	Std. Error	t-value
(Intercept)	0.106	0.315	0.337*
Linear poly	-0.857	0.237	-3.616*
Quadratic poly	-1.799	0.246	-7.311*
Cubic poly	0.966	0.249	3.887*
NP2	0.241	0.023	10.485*
High	0.201	0.023	8.761*
Linear:NP2	1.152	0.237	4.858*
Quadratic:NP2	-0.2	0.246	-0.814
Cubic:NP2	0.191	0.249	0.769
Linear:High	0.774	0.237	3.263*
Quadratic:High	-0.363	0.246	-1.473
Cubic:High	-0.312	0.249	-1.255
NP2:High	0.001	0.023	0.065
Linear:NP2:High	-0.578	0.237	-2.438*
Quad:NP2:High	0.342	0.246	1.390
Cubic:NP2:High	0.882	0.249	3.547*

Table 2: Growth curve analysis in a linear mixed effects regression model. The \* indicates a significant effect at the  $\alpha = .05$  threshold.

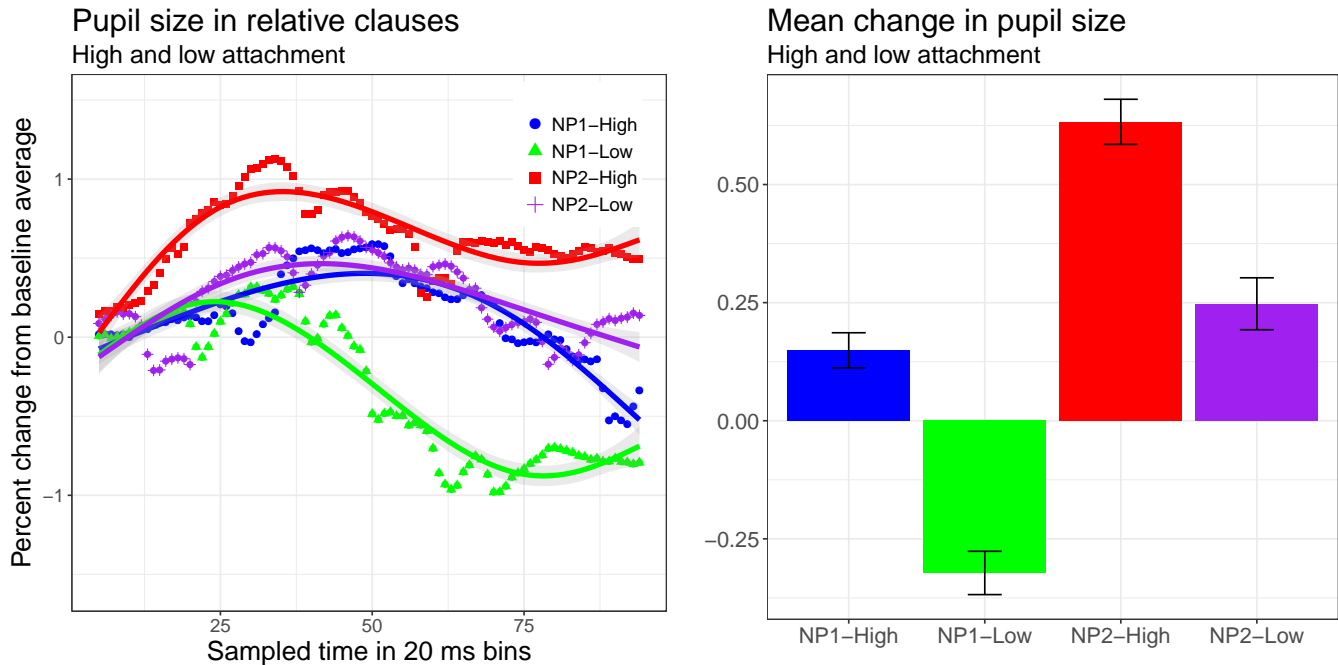


Figure 1: Percent change of pupil dilation over time with standard errors (left panel) and mean percent change of pupil size collapsing across time bins (right panel).

In growth curve analyses, polynomial terms capture distinct components of the functional form of a curve as it develops over time, and will be interpreted with respect to pupil dilation as follows (Kuchinsky et al., 2013; McGarigle et al., 2017). The INTERCEPT corresponds to the overall mean pupillary response, so that positive coefficients indicate greater amplitudes. The LINEAR polynomial term coefficient corresponds to the slope of pupillary response, so that a positive increase in the coefficient indicates more steeply rising pupil dilation. The QUADRATIC polynomial term describes the shape of the primary inflection point, revealing the degree to which the curve is non-linear. Negative quadratic coefficients indicate an inverted U-shaped curve, characteristic of pupil peaks. The CUBIC term captures the properties of any secondary inflection point in the curve, so that positive coefficients indicate that pupil dilation peaks are more compressed or transient, rising and falling more sharply.

Positive coefficients of High attachment and NP2 boundary indicate increased average pupil dilation (the area under the curve) for High attachment over Low attachment, and NP2 boundary location over NP1 boundary location, respectively. Interactions between the planned predictor variables and the polynomial terms indicate how the experimental conditions differentially affect the shape of the pupillary response over time.

The mean change in pupil size for each condition is shown in Figure 1. In the left panel, the shape of the best-fitting non-linear regression line is plotted against change in pupil size within a 2000ms period immediately after the relative clause. The values on the vertical axis represent the percent

change from the baseline average, defined here as the entire 100ms segment of silence inserted between the end of the relative clause and the remainder of the sentence (*although the club was really noisy*). The right panel reports the overall mean pupil change (with standard errors in grey) in pupil size during the same period for visual comparison.

In the growth curve model, effects of all three orthogonal polynomial terms were observed. Modulo the manipulation, pupil growth was less steep (a negative Linear coefficient), showed greater inverted U-shaped curve (a negative Quadratic coefficient), and a sharper secondary point of inflection (a positive Cubic coefficient), corresponding to a change or bend in the direction of the response.

More importantly, the two experimental factors in the manipulation showed that a prosodic boundary after NP2, and relative clauses that were grammatically disambiguated to a High attachment relation elicited greater pupil excursions from the grand mean compared to their respective NP1 and Low attachment reference levels. Both NP2 boundary and High attachment conditions also elicited more steeply rising slopes, as indicated by their interaction with the linear orthogonal polynomial.

In addition, the interaction between NP2 boundary and High attachment conditions further interacted with linear and cubic polynomials, indicating that the NP2-High condition elicited a smaller slope and increased transience of the pupil peak. The overall interaction is perhaps best visualized in the right panel of Figure 1, where the effect of NP2 is greater for High attachment conditions.

Perhaps more intuitively, the plot in the left panel of Fig-

ure 1 suggests the following conclusions. First, the conditions where the syntax and the prosody aligned largely conform to expectations. The theoretical baseline condition (NP1-Low) elicited the least extreme growth in pupil size, whereas the NP2-High elicited the most extreme changes. Low attachment is thought to be compatible with a boundary after NP1. Low attachment instantiates the empirically preferred relation between a RC and a complex nominal head in English, and, by hypothesis, is the least taxing relation to compute. The fact that the response was relatively muted in this condition is therefore entirely compatible with current linguistic theory. Similarly for the NP2-High condition: a boundary after NP2 aligns with the proposed syntax of High attachment structures. Assuming that non-local relations, including High attachment, are costly to compute, the fact that the NP2-High condition elicited the most extreme response is also consistent with current theory. However, our findings do not support the possibility that the bias against High attachment could be solely attributed to lack of supporting prosodic information; even when structures were disambiguated by prosodic boundary location, High attachment structures elicited increased cognitive load.

Second, the conditions where the prosodic grouping did not align with the syntactic constituency reveal a more complicated pattern. We predicted that the NP1-High attachment condition would be more anomalous than the NP2-Low attachment condition. Our reasoning was that the prosody of NP1-High would encourage grouping the relative clause and NP2 together, e.g. *the musicians who was really quiet*, creating a local number mismatch violation between NP2 (*the musicians*) and the verbal agreement marker (*was*). In contrast, the NP2-Low condition (*the musician % who was really quiet*) is locally grammatically coherent despite an infelicitous prosodic boundary. However, the two mismatching conditions elicited similar response patterns.

We entertained three main possible explanations. The first was that, in cases of mismatching cues, the processor makes weaker online processing commitments, and may defer the attachment decision until later, if it commits at all, as in models employing syntactic underspecification for attachment ambiguities (e.g., Frazier & Clifton, 1996). This interpretation is broadly compatible with results from Johnson et al.'s (2014) digit span task, which found decreased pupil size in response to digit sequences exceeding normal capacity. Decreased pupillary response may also indicate that the subject has abandoned an excessively difficult task, highlighting the role of attention in relating pupil size growth to cognitive effort (see also Beatty, 1982 and Winn et al., 2018)

A second possibility was that systematic differences between items may have prompted different processing strategies. For example, half of the items were disambiguated with the singular auxiliary marker (*was*), half were disambiguated with the plural marker (*were*). Our intuition was that relative clause attachment relations that were disambiguated with a singular form (*Everybody met the brother % of the musicians*

*who was really quiet*) would be less anomalous than cases with plural disagreement (*I got a call from the friends % of the lawyer who were in politics*). We further addressed this possibility by including which number (singular vs. plural) was used to disambiguate the attachment location. Impressionistically, NP1-High conditions elicited greater pupil excursions in the Plural condition. However, grammatical number failed to differentiate effects within the statistical model.

A third explanation was that the processor resolves to High or Low attachment on the basis of another unidentified factor, such as by-item differences in boundary strength, prominence, or plausibility. To address these possibilities, we conducted a *post-hoc* boundary identification and rating norming study. The post RC material was removed, and the items were placed into four counterbalanced lists along with 46 filler items from the pupillometry experiment. Twenty additional participants from the same population as before were instructed to listen to each sentence over headphones in a noise-attenuated sound booth as many times as necessary, in order to manually mark prosodic boundaries on written versions of the sentence, and to rate how well the produced sentence matched its likely intended meaning (1 = completely unnatural, 7 = completely natural).

While participants were at ceiling (99.8%) at identifying the prosodic boundaries at the intended location after NP1 and NP2, additional boundaries after NP2 were perceived in post-NP1 boundary conditions 22% of the time. Subjects may have reverted to their default prosodic preference for an additional, potentially weaker, boundary before the RC (as discussed in Jun, 2003). Consistent with that interpretation, there was a main effect of prosody in ratings ( $p < .01$ ), in which NP2 conditions ( $M=4.80$ ,  $SE=0.12$ ) were rated a more natural match with the intended meaning than NP1 conditions ( $M=3.66$ ,  $SE=0.11$ ). The penalty for non-local relations was evident in the ratings, as well. Sentences with Low attachment RCs ( $M=4.41$ ,  $SE=0.12$ ) were rated as more naturally matching with the prosody than sentences with High attachment RCs ( $M=4.05$ ,  $SE=0.12$ ),  $p < .01$ . The two factors did not significantly interact, suggesting that subjects were not sensitive to mismatches between syntactic and prosodic cues in this relatively conscious offline task.

## General discussion

We used pupillometry to explore how prosodic and structural information interact during online language comprehension. Though relatively under-utilized in language processing research, pupillometry offers a promising methodological avenue for exploring how prosodic and structural information are integrated in real time processing. This method is especially useful for investigating how listeners use acoustic information to construct an interpretation, and offers a naturalistic and cost-effective complement to better known methods, such as neuroimaging or eye tracking in the visual world paradigm.

The results of the study replicate the well-studied bias for Low attachment of relative clauses in complex noun phrases

in English, a preference known to be modulated by prosodic boundary placement. However, our results cast doubt on an account which would attribute the preference to a lack of prosodic information alone. Even in the presence of overt prosodic boundaries, sentences with non-local dependencies were found to elicit online processing penalties. In addition, when the prosodic and grammatical information did not agree, pupillary response was reduced, indicating that prosody and structure are incorporated into an unfolding representation in concert. While more work is needed to investigate how language users integrate multiple sources of information together, the current study is compatible with the claim that comprehenders may avoid or delay making certain processing decisions in the face of inconsistent information.

### Acknowledgments

Our thanks to Bethany Sturman for recording and pre-processing the materials, and to Joonhwa Kim, Nathan Mallipeddi, Alison Suh, Chenchen Wang, and Rebecca Wu for administering the study. Thanks also to Sun-Ah Jun for discussion of prosodic boundaries and to the UCLA Psycholinguistics Seminar for insightful comments.

### References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292.
- Beatty, J., & Kahneman, D. (1966). Pupillary changes in two memory tasks. *Psychonomic Science*, *5*, 371–372.
- Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.
- Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*, 73–105.
- Demberg, V., & Sayeed, A. (2016). The frequency of rapid pupil dilations as a measure of linguistic processing difficulty. *PloS One*, *11*, e0146194.
- Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *The Quarterly Journal of Experimental Psychology*, *63*, 639–645.
- Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam, The Netherlands: John Benjamins Publishing.
- Fernández, E. M., & Sekerina, I. A. (2015). The interplay of visual and prosodic information in the attachment preferences of semantically shallow relative clauses. In L. Frazier & E. Gibson (Eds.), *Explicit and Implicit Prosody in Sentence Processing*. Springer.
- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, *27*, 285–319.
- Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.
- Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291–325.
- Fromont, L. A., Soto-Faraco, S., & Biau, E. (2017). Searching high and low: Prosodic breaks disambiguate relative clauses. *Frontiers in Psychology*, *8*, 96.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.
- Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition*, *59*, 23–59.
- Gilboy, E., Sopena, J.-M., Clifton, C., Jr., & Frazier, L. (1995). Argument structure and association preferences in Spanish and English complex NPs. *Cognition*, *54*, 131–167.
- Grillo, N., Costa, J., Fernandes, B., & Santi, A. (2015). Highs and lows in English attachment. *Cognition*, *144*, 116–122.
- Grillo, N., & Costa, J. a. (2014). A novel argument for the universality of parsing principles. *Cognition*, *133*, 156–187.
- Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: The two sides of relative clause attachment. In M. W. Crocker, M. J. Pickering, & C. Clifton Jr. (Eds.), *Architectures and Mechanisms for Language Processing*. Cambridge, UK: Cambridge University Press.
- Hess, E. H., & Polt, J. M. (1960). Pupil size as related to interest value of visual stimuli. *Science*, *132*, 349–350.
- Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L., & Bunge, S. A. (2014). Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Frontiers in psychology*, *5*, 218.
- Jun, S.-A. (2003). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research*, *32*, 219–249.
- Just, M., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*, 310–339.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*, 1583–1585.
- Kim, J. H., & Christianson, K. (2013). Sentence complexity and working memory effects in ambiguity resolution. *Journal of psycholinguistic research*, *42*, 393–411.
- Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47.
- Kuchinke, L., Võ, M. L.-H., Hofmann, M., & Jacobs, A. M. (2007). Pupillary responses during lexical decisions vary with word frequency but not emotional valence. *International Journal of Psychophysiology*, *65*, 132–140.

- Kuchinsky, S. E., Ahlstrom, J. B., Vaden Jr, K. I., Cute, S. L., Humes, L. E., Dubno, J. R., & Eckert, M. A. (2013). Pupil size varies with word listening and response selection difficulty in older adults with hearing loss. *Psychophysiology*, *50*, 23–34.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.
- Laeng, B., Sirois, S., & Gredebäck, G. (2012). Pupillometry: A window to the preconscious? *Perspectives on Psychological Science*, *7*, 18–27.
- Lund, K., & Burgess, C. (1996). Hyperspace analogue to language (HAL): A general model semantic representation. *Brain and Cognition*, *30*, 203–208.
- McGarrigle, R., Dawes, P., Stewart, A. J., Kuchinsky, S. E., & Munro, K. J. (2017). Pupillometry reveals changes in physiological arousal during a sustained listening task. *Psychophysiology*, *54*, 193–203.
- Mirman, D. (2016). *Growth curve analysis and visualization using R*. CRC Press.
- R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Scheepers, C., Mohr, S., Fischer, M. H., & Roberts, A. M. (2013). Listening to limericks: a pupillometry investigation of perceivers' expectancy. *PLoS One*, *8*, e74986.
- Schmidtke, J. (2018). Pupillometry in linguistic research: An introduction and review for second language researchers. *Studies in Second Language Acquisition*, *40*, 529–549.
- Wilhelm, B., Wilhelm, H., & Lüdtke, H. (1999). Pupillography: Principles and applications in basic and clinical research. *Pupillography: Principles, methods and applications*, 1–11.
- Winn, M. B., Wendt, D., Koelewijn, T., & Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: An introduction for those who want to get started. *Trends in Hearing*, *22*, 1–32.
- Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*, 133–141.

## Sample stimuli from experiment

Six additional experimental items (from a total of 20 sentences). Low attachment disambiguation is presented prior to High attachment disambiguation. Disambiguation was evenly balanced across singular (*was*) and plural (*were*) auxiliary markers.

1. I got a call from the friends (%) of the lawyer (%) who was / were volunteering for the campaign // but my phone died halfway through the call.
2. We were all listening to the neighbor (%) of the pilots (%) who were / was raising exotic pets // even though we were in a hurry.
3. I spoke to the apprentices (%) to the librarian (%) who was / were wearing blue jeans // but I do not remember what we discussed.
4. Somebody saw the manager (%) of the architects (%) who were / was planning to buy a Mercedes // although it was very dark outside.
5. Someone kissed the sisters (%) of the medic (%) who was / were expecting to work late // yet nobody saw it happen.
6. Everybody admired the parent (%) of the artists (%) who were / was dancing the waltz // even though there was no music playing.